Transportation Distance and the Central Limit Theorem

Svetlana Ekisheva* and Christian Houdré†

February 2, 2008

Abstract

For probability measures on a complete separable metric space, we present sufficient conditions for the existence of a solution to the Kantorovich transportation problem. We also obtain sufficient conditions (which sometimes also become necessary) for the convergence, in transportation, of probability measures when the cost function is continuous, non-decreasing and depends on the distance. As an application, the CLT in the transportation distance is proved for independent and some dependent stationary sequences.

Keywords: Kantorovich transportation problem, convergence in transportation distance, Central Limit Theorem in transportation distance, Wasserstein distance, strong mixing sequences, associated sequences.

AMS classifications: 60B05, 60F05, 60F25, 28A33.

1 Introduction

Let (M, d) be a metric space and let $c : M \times M \to \mathbf{R}$, be a non-negative Borel function. The transportation c-distance $T_c(\mu, \nu)$ between two probability measures μ and ν defined on the Borel σ-field $\mathcal{B}(M)$ is given via

$$T_c(\mu, \nu) = \inf \mathbf{E}c(X, Y).$$

Above, the infimum is taken over all M-valued random elements X and Y defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and having, respectively, μ and ν for probability distribution. In other words,

$$T_c(\mu, \nu) = \inf_{\Pi} \int c(x, y) d\pi(x, y), \tag{1}$$

where the infimum is taken over the set Π of all probability measures on $\mathcal{B}(M \times M)$ with marginals μ and ν. The transportation distance is related to the celebrated Kantorovich transportation problem: if μ and ν are two distributions of mass and if $c(x, y)$ represents the cost of transporting a unit of mass from the location x to the location y, what is the minimal total transportation cost to transfer μ to ν? The minimal total transportation cost is exactly the transportation distance corresponding to the cost function c.

*Schools of Biology and †Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0230 USA sveta.ekisheva@bme.gatech.edu; houdre@math.gatech.edu

The c-transportation distance with $c(x,y) = d^p(x,y)$, $p \geq 1$, is associated to the Wasserstein or Mallows p-distance W_p, $W_p(\mu,\nu) = (T_{d^p}(\mu,\nu))^{1/p}$. If M is the real line $\mathbf{R}$ with the Euclidean distance, the Wasserstein-Mallows p-distance between two distribution functions F and G has the following useful representation

$$W_p^p(F,G) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt, \tag{2}$$

where the inverse transformation of F is defined as

$$F^{-1}(t) = \sup\{x \in \mathbf{R} : F(x) \leq t\}.$$

The representation (2) was obtained when $p = 1$ by Salvemini [20] (for discrete distributions) and by Dall'Aglio [7] (in the general case), while for $p = 2$ it is due to Mallows [15]. It implies that the random variables $X = F^{-1}(U)$ and $Y = G^{-1}(U)$, where U is a uniform random variable on $(0,1)$, are minimizers of the total transportation cost in the transportation problem. Major [14] generalized (2) to a convex cost function $c(x,y) = c(x-y)$:

$$T_c(F,G) = \int_0^1 c(F^{-1}(t) - G^{-1}(t))dt.$$

The representation (2) is an important tool in proving the following convergence result.

Let $p = 1, 2$ and let F_n, F be distribution functions on $\mathbf{R}$ such that for any n, $\int |x|^p dF_n < +\infty$, and $\int |x|^p dF < +\infty$. Then

$$W_p(F_n, F) \to 0 \iff \begin{cases} (a)\ F_n \Longrightarrow F, \\ (b) \int |x|^p dF_n \to \int |x|^p dF. \end{cases} \tag{3}$$

For $p = 1$ the equivalence (3) was proved by Dobrushin [9], while for $p = 2$ it is due to Mallows [15].

Bickel and Freedman [2] extended the statement (3) to probability measures μ_n and μ defined on a separable Banach space $(\mathbf{B}, \|\cdot\|)$ and to all $p \in [1, +\infty)$ as follows:

Let $1 \leq p < \infty$, and let $\int \|x\|^p \mu_n(dx) < \infty$, $\int \|x\|^p \mu(dx) < \infty$. Then $W_p(\mu_n, \mu) \to 0$ as $n \to \infty$ is equivalent to each of the following.

(a) $\mu_n \Longrightarrow \mu$ and $\int \|x\|^p \mu_n(dx) \to \int \|x\|^p \mu(dx)$.

(b) $\mu_n \Longrightarrow \mu$ and $\|\cdot\|^p$ is uniformly μ_n-integrable.

(c) $\int \phi(x)\mu_n(dx) \to \int \phi(x)\mu(dx)$ for every continuous ϕ such that $\phi(x) = O(\|x\|^p)$ at infinity.

Since in general an analog of the representation (2) does not exist for probability measures on a Banach space, Bickel and Freedman proved, in their setting, the existence of a solution to the transportation problem for $c(x,y) = \|x-y\|^p$.

Recently, Ambrosio, Gigli, and Savaré proved ([1], Proposition 7.1.5) an analog of part (b) of the above result for probability measures on a Radon space X (see also Lemma 5.1.7 and Remark 7.1.11 there). These authors also established the existence of a solution to the Kantorovich transportation problem in X for a wide class of cost functions. We use this existence result to prove criteria for the convergence in T_c with $c(x,y) = C(d(x,y))$, where C is a non-decreasing continuous function satisfying the doubling condition (6) which controls

the rate of growth of C (Theorem 2 and Corollary 1). Since the class of such cost functions includes all the d^ps, $p \geq 1$, the convergence results of Bickel and Freedman as well as those of Ambrosio, Gigli, and Savaré follow from Corollary 1. Note that instead of the Radon space X (a separable metric space, where, by definition, every probability measure is tight), we consider more familiar in the theory of probability object, a complete separable space (M, d) where completeness and separability together provide the tightness of a probability measure; all our arguments remain true for a Radon space (see also Remark 1).

In Theorem 2 we also obtain sufficient conditions for the convergence of probability measures in the transportation distance without assuming the doubling condition on C. For instance, any convex $C : \mathbf{R}^+ \to \mathbf{R}^+$ with $C(0) = 0$ satisfies Theorem 2. We then provide an example of a C growing exponentially fast for which the converse implication does not hold.

2 Convergence in Transportation Distance

The following result of Ambrosio, Gigli, and Savaré [1] asserts the existence in Π of a probability measure which minimizes the total transportation cost under rather weak assumptions on the cost function. For the sake of completeness, we include a self-contained proof in Section 4.

Theorem 1. *Let (M, d) be a complete separable metric space, and let $T_c(\mu, \nu)$ be defined by (1) with $c : M \times M \to [0, +\infty)$ lower semicontinuous.*

Then there exists $\pi^* \in \Pi$ such that $\int c(x, y) d\pi^*(x, y) = T_c(\mu, \nu)$. Or, equivalently, there exists a pair of random elements X and Y with respective distributions μ and ν, such that $\mathbf{E}c(X, Y) = T_c(\mu, \nu)$.

Remark 1. In the corresponding statement in [1] the space (M, d) need not be a complete separable metric space but just a Radon space. In fact, our proof also shows that completeness is unnecessary and that the tightness of μ and ν will suffice. On the other hand, the hypothesis of separability of (M, d) can be weakened to the topological separability if both μ and ν have separable supports (see Billingsley [3], Appendix III).

The Kantorovich problem is closely related to the Monge transportation problem which is the problem of finding a map s^* pushing μ forward to ν (i.e. such that $\nu(B) = \mu(s^{-1}(B))$ for any Borel set B) and minimizing the total transportation cost: $\inf_s \int c(x, s(x)) d\mu = \int c(x, s^*(x)) d\mu$, where the infimum is taken over all Borel maps s pushing μ forward to ν. A solution s^* to the Monge transportation problem uniquely determines a probability measure π^* on $M \times M$ such that the random elements X and Y, $Y = s^*(X)$, with respective distributions μ and ν have joint law π^*. This measure π^* minimizes the Monge transportation cost:

$$\int c(x, y) d\pi^*(x, y) = \inf_{\Pi^*} \int c(x, y) d\pi(x, y), \tag{4}$$

where the infimum is taken over the set Π^* of joint distributions of M-valued random elements X and Y with respective distributions μ and ν and such that Y is measurable with respect to the Borel field $\sigma(X)$. Comparing (4) and (1) yields the relation $\Pi^* \subset \Pi$, which immediately leads to the following conclusions: (i) the (Kantorovich) transportation distance $T_c(\mu, \nu)$

never exceeds the Monge transportation distance $\tilde{T}_c(\mu,\nu)$,

$$\tilde{T}_c(\mu,\nu) = \inf_{\Pi^*} \int c(x,y)d\pi(x,y) = \inf_s \int c(x,s(x))d\mu;$$

(ii) a probability measure π^* corresponding to the solution s^* of the Monge transportation problem (MTP) is not necessarily a solution to the Kantorovich transportation problem (KTP); conversely, a solution π' of the KTP, where π' is the joint distribution of X and Y, is a solution to the MTP if and only if there exists a Borel map s' such that $Y = s'(X)$.

For random elements X and Y in a Hilbert space Cuesta and Matran [6] have provided conditions for the existence of an increasing map s, $s(X) = Y$, such that $W_2^2(\mu,\nu) = \mathbf{E}\|X - s(X)\|^2$, i.e. X and $Y = s(X)$ give the solution to both the MTP and the KTP. They also showed that if μ is either absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}^k$ or is a Gaussian measure on a Hilbert space, then these conditions are satisfied. For compactly supported absolutely continuous distributions on $\mathbf{R}^k$ and a convex cost function $c(x-y)$ Caffarelli [5] has determined the form of the optimal map (the solution to the MTP) as a gradient of c; the uniqueness of the solution is also obtained there. Simultaneously, Gangbo and McCann [11] proved the same results for non-necessarily boundedly supported probability measures. They also showed that the solution to the MTP is the KTP solution as well, and that a similar result holds true for $c(x,y) = l(\|x-y\|)$, where l is strictly concave. Note that in all the existence statements mentioned above, the conditions of Theorem 1 are satisfied. A comprehensive review of the results on the solutions to the KTP and the MTP can be found in the books of Rachev and Rüschendorf [19].

The main result of the work presented here is now given.

Theorem 2. *Let μ_n and μ be probability measures on a complete separable metric space (M,d) and let $c : M\times M \to \mathbf{R}$ be such that $c(x,y) = C(d(x,y))$, where $C : [0,+\infty) \to [0,+\infty)$ is a non-decreasing continuous function with $C(0) = 0$. Let*

$$\int C(2d(x,a))\mu_n(dx) < \infty, \qquad \int C(2d(x,a))\mu(dx) < \infty \tag{5}$$

for some (and, therefore, for all) $a \in M$. Then

$$\left.\begin{array}{l} (a)\ \mu_n \Longrightarrow \mu, \\ (b) \int C(2d(x,a))\mu_n(dx) \to \int C(2d(x,a))\mu(dx) \end{array}\right\} \Longrightarrow T_c(\mu_n,\mu) \to 0.$$

Conversely, if $T_c(\mu_n,\mu) \to 0$, then $\mu_n \Longrightarrow \mu$. If, additionally, C satisfies a doubling condition, i.e. if there exists a positive constant λ such that for all $y \geq 0$

$$C(2y) \leq \lambda C(y), \tag{6}$$

then

$$T_c(\mu_n,\mu) \to 0 \Longleftrightarrow \begin{cases} (a)\ \mu_n \Longrightarrow \mu, \\ (b) \int C(2d(x,a))\mu_n(dx) \to \int C(2d(x,a))\mu(dx). \end{cases}$$

Corollary 1. *If, in the setting of Theorem 2, C satisfies (6), then*

$$(6) \Longleftrightarrow \int C(d(x,a))\mu_n(dx) < \infty, \int C(d(x,a))\mu(dx) < \infty,$$

and thus

$$T_c(\mu_n,\mu) \to 0 \Longleftrightarrow \begin{cases} (a)\ \mu_n \Longrightarrow \mu, \\ (b')\int C(d(x,a))\mu_n(dx) \to \int C(d(x,a))\mu(dx). \end{cases}$$

Corollary 1 is equivalent to a result of Rachev (Theorem 1 in [18]) proved by using the relations between the Lévy-Prokhorov metric and the T_c-distance. Since for any $p \geq 1$, the function $c(x,y) = d^p(x,y)$ satisfies the conditions of Theorem 2 as well as (6) with $\lambda = 2^p$, Corollary 1 recovers part (a) in the result of Bickel and Freedman mentioned above. Ambrosio, Gigli, and Savaré [1] proved an analog of Theorem 2 in a Hilbert space when cost function is continuous, strictly increasing and surjective map.

Note that the class of functions C covered by Theorem 2 includes functions with a faster than polynomial rate of growth at infinity (e.g. $C(d(x,y)) = \exp(d(x,y)) - 1$). For functions C growing exponentially fast at infinity, and in contrast to $C(d(x,y)) = d^p(x,y)$, $T_c(\mu_n,\mu) \to 0$ need not imply the convergence of $\int C(2d(x,a))\mu_n(dx)$ to $\int C(2d(x,a))\mu(dx)$, for some $a \in M$. Indeed, one can take the probability measures μ_n and μ on $\mathbf{R}$ defined in Example 1, below, and $c(x,y) = C(|x-y|) = \exp(|x-y|) - 1$.

As a corollary to Theorem 2 we obtain the following result relating the convergence in total variation to the convergence in transportation distance. It is well known that the total variation distance itself is a particular case of transportation distance (with $c(x,y) = 2\mathbf{1}_{\{x \neq y\}}$).

Corollary 2. *Let μ and ν be boundedly supported probability measures on a complete separable metric space (M,d), and let ϕ be a continuous function on (M,d). Then*

$$\left|\int \phi(x)\mu(dx) - \int \phi(x)\nu(dx)\right| \leq L_\phi \|\mu - \nu\|_{TV}$$

for some positive constant L_ϕ.

Let μ_n be probability measures on M with respective supports K_n, $n \geq 1$. Let $\cup_n K_n$ be bounded. If $c(x,y) = C(d(x,y))$, where $C : [0,+\infty) \to [0,+\infty)$ is non-decreasing, continuous with $C(0) = 0$, then

$$\|\mu_n - \mu\|_{TV} \to 0 \Rightarrow T_c(\mu_n,\mu) \to 0.$$

Without the boundedness restriction on $\cup K_n$ the last implication is not true, as the following example shows.

Example 1. Let μ be the uniform distribution on $(0,1)$ and, for all $n \in \mathbf{N}$, let

$$\mu_n(dx) = \frac{n-1}{n}\mathbf{1}_{(0,1)}(x)dx + \frac{1}{n}\delta_{x_n}(dx),$$

$x_n \notin (0,1)$. Then

$$\|\mu_n - \mu\|_{TV} = \int_0^1 |f_n(x) - 1|dx + \mu_n(x_n) = \frac{2}{n} \to 0$$

as $n \to \infty$. Hence $\mu_n \xrightarrow{TV} \mu$ for any choice of the sequence (x_n). Let $c(x, y) = C(|x - y|)$, with $C : [0, +\infty) \to [0, +\infty)$, $C(0) = 0$, convex, also satisfying (6). Then,

$$\int C(|x|)\mu(dx) = \int_0^1 C(|x|)dx \leq \max_{0<|x|<1} C(|x|) < +\infty,$$

$$\int C(|x|)\mu_n(dx) = \int_0^1 \frac{n-1}{n} C(|x|)dx + \mu_n(x_n)C(|x_n|) \leq \max_{0<|x|<1} C(|x|)\frac{n-1}{n} + \frac{C(|x_n|)}{n} < +\infty,$$

for any n. So all the conditions of Corollary 1 are satisfied. Since weak convergence is implied by convergence in total variation, $T_c(\mu_n, \mu) \to 0$ holds if and only if $\int C(|x|)\mu_n(dx) \to \int C(|x|)\mu(dx)$. Take $x_n = 2^n$, then $C(|x_n|) = C(2^n) \geq 2^{n-1}C(2)$ and $C(|x_n|)/n \to +\infty$ as $n \to \infty$. Therefore,

$$\int C(|x|)\mu_n(dx) \geq \frac{C(|x_n|)}{n} \to +\infty \neq \int C(|x|)\mu(dx).$$

By Corollary 1, $T_c(\mu_n, \mu)$ does not converge to 0.

3 Applications to the Central Limit Theorem

Next, we apply Theorem 2 to obtain the CLT in the transportation distance. We provide sufficient conditions for the convergence of the laws of the normalized sums to the standard Gaussian measure on $\mathbf{R}$ for stationary sequences which are either independent, strongly mixing or associated.

3.1 Independent sequences

Let (X_n) be a sequence of independent identically distributed random variables, $\mathbf{E}X_1 = 0$, $\mathbf{E}X_1^2 = \sigma^2$, $0 < \sigma < +\infty$. Let $S_n = \sum_{k=1}^n X_k$. Then by the classical central limit theorem

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0, 1).$$

Let μ_n denote the probability law of the normalized sum $S_n/(\sigma\sqrt{n})$, and let γ be the standard Gaussian measure on $\mathbf{R}$. We find additional conditions on the sequence (X_n) and on the cost function to obtain the convergence of μ_n to γ in the T_c-distance.

Theorem 3. *Let $c(x, y) = C(|x - y|)$, where $C : [0, +\infty) \to [0, +\infty)$ is a non-decreasing continuous function with $C(0) = 0$.*

(i) *If there exists $p \geq 2$ such that $C(x) = O(x^p)$ at infinity and $\mathbf{E}|X_1|^p < +\infty$, then $T_c(\mu_n, \gamma) \to 0$.*

(ii) *Otherwise, let $\mathbf{E}C(4\sqrt{2}|Z|) < +\infty$ and let $\sum_{k=1}^{\infty} k^k \mathbf{E}X_1^{2k} < +\infty$, then $T_c(\mu_n, \gamma) \to 0$.*

The CLT in the W_2-distance was proved by Tanaka [21] for distributions on $\mathbf{R}$ and by Cuesta and Matran [6] for distributions on a Hilbert space; both results require the finiteness of the fourth moment. Very recently, Johnson and Samworth [12], [13] proved that $W_p(\mu_n, \gamma) \to 0$, $p \geq 2$, under the condition $\mathbf{E}|X_1|^p < \infty$. This statement is a particular case of part (i) of Theorem 3. However, these authors also proved the convergence to an α-stable law in the Wasserstein-Mallows α-distance, $\alpha \in (0, 2)$.

We will prove Theorem 3 by applying Theorem 2. The CLT yields weak convergence; therefore, to prove convergence in T_c-distance, we need to verify the convergence of $\int C(2|x|)d\mu_n$ to $\int C(2|x|)d\gamma$. To do so, we use Rosenthal's inequality which asserts that for stationary independent sequence (X_n) of centered random variables

$$\mathbf{E}|S_n|^p \leq K(p) \left(n\mathbf{E}|X_1|^p + n^{p/2}(\mathbf{E}|X_1|^2)^{p/2} \right) \tag{7}$$

for $p > 1$ and a positive constant $K(p)$ depending only on p (Petrov [17]).

3.2 Strong mixing sequences

The coefficients α_n of strong mixing of a random sequence (X_n) are defined as

$$\alpha_n = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_1^k,\ B \in \mathcal{F}_{k+n}^{+\infty}, k \geq 1\},$$

where $\mathcal{F}_k^{k+m}$ is the σ-field generated by the random variables $X_k, X_{k+1}, ..., X_{k+m}$. A sequence is said to satisfy a strong mixing condition if $\alpha_n \to 0$ as $n \to +\infty$.

A CLT for a stationary strong mixing sequence (X_n) was proved by Denker [8] in the following form. Let $\mathbf{E}X_1 = 0$, $\mathbf{E}X_1^2 = \sigma^2$, $0 < \sigma < +\infty$, and $\sigma_n^2 = \mathbf{E}S_n^2 = nh(n)$, where $h(n)$ is a slowly varying function. Let $S_n = \sum_{k=1}^n X_k$, and let μ_n be the law of S_n/σ_n. Then $\mu_n \Longrightarrow \gamma$, where γ is the standard Gaussian measure on $\mathbf{R}$.

To obtain the convergence of μ_n to γ in T_c, we need additional conditions on the rate of decay of the coefficient α_n providing a moment inequality for sums. Such a result exists, it is due to Yokoyama [22] and asserts that if (X_n) is a stationary strong mixing sequence such that $\mathbf{E}X_1 = 0$, $\mathbf{E}|X_1|^{p+\delta} < +\infty$, $p > 2$, $\delta > 0$ and

$$\sum_{n=1}^{\infty} (n+1)^{\frac{p}{2}-1} (\alpha_n)^{\frac{\delta}{p+\delta}} < +\infty, \tag{8}$$

then

$$\mathbf{E}|S_n|^p \leq K(p) n^{\frac{p}{2}}, \tag{9}$$

where the positive constant $K(p)$ depends only on p.

Theorem 4. *Let $c(x,y) = C(|x-y|)$, where $C : [0,+\infty) \to [0,+\infty)$ is a non-decreasing continuous function with $C(0) = 0$.*

(i) *If there exist $p > 2$ and $\delta > 0$ such that the condition* (8) *is satisfied and $C(x) = O(x^p)$ at infinity, then $T_c(\mu_n, \gamma) \to 0$.*

(ii) *Otherwise, let $\mathbf{E}C(4\sqrt{2}|Z|) < +\infty$, let $\sum_{k=1}^{\infty} k^k \mathbf{E}X_1^{2k} < +\infty$, and let* (8) *be satisfied for all $p > 2$, then $T_c(\mu_n, \gamma) \to 0$.*

To prove Theorem 4, we once again apply Theorem 2. Since the corresponding CLT implies weak convergence to the Gaussian measure, it is sufficient to show the convergence of $\mathbf{E}C(2|Y_n|)$ to $\mathbf{E}C(2|Z|)$.

3.3 Associated sequence

Recall that a set of random variables $\xi = (\xi_1, ..., \xi_m)$ is called associated if for any two coordinatewise increasing functions $f,\ g : \mathbf{R}^m \to \mathbf{R}$ such that $\mathbf{E}f(\xi)g(\xi)$, $\mathbf{E}f(\xi)$ and $\mathbf{E}g(\xi)$ exist,

$$\mathrm{Cov}\,(f(\xi), g(\xi)) \geq 0.$$

An infinite set of random variables is associated if all of its finite subsets are associated.

Newman [16] proved the CLT for associated sequence under the following conditions. Let (X_n) be a stationary associated sequence, $\mathbf{E}X_1 = 0$, $\mathbf{E}X_1^2 = \sigma^2$, $0 < \sigma < +\infty$, $\sigma_n^2 = \mathbf{E}S_n^2 = nh(n)$, where $h(n)$ is a slowly varying function. Let μ_n be the law of S_n/σ_n and let γ be the standard Gaussian measure on $\mathbf{R}$. Then $\mu_n \Longrightarrow \gamma$.

Asymptotic independence for associated sequence (X_n) is usually stated in terms of the Cox-Grimmett coefficient $u(n)$ defined by:

$$u(n) = \sup_{k\geq 1} \sum_{j:|j-k|\geq n} \mathrm{Cov}(X_j, X_k).$$

For a stationary sequence the Cox-Grimmett coefficient is just the tail of the series of covariances:

$$u(n) = 2 \sum_{k=n+1}^{\infty} \mathrm{Cov}(X_1, X_k).$$

To prove the convergence of μ_n to γ in the transportation distance, we use a condition on the rate of decay of the Cox-Grimmett coefficient. This condition implies the following moment inequality for sums (Birkel [4]). If (X_n) is a stationary associated sequence, $\mathbf{E}X_1 = 0$, $\mathbf{E}|X_1|^{p+\delta} < +\infty$, $p > 2$, $\delta > 0$ and

$$u(n) \leq Bn^{-\frac{(p-2)(p+\delta)}{2\delta}}, \tag{10}$$

then

$$\mathbf{E}|S_n|^p \leq K(p)n^{\frac{p}{2}}, \tag{11}$$

where the positive constant $K(p)$ depends only on p.

Theorem 5. *Let $c(x,y) = C(|x-y|)$, where $C : [0,+\infty) \to [0,+\infty)$ is a non-decreasing continuous function with $C(0) = 0$.*

(i) *If there exist $p > 2$ and $\delta > 0$ such that the condition* (10) *is satisfied and $C(x) = O(x^p)$ at infinity, then $T_c(\mu_n, \gamma) \to 0$.*

(ii) *Otherwise, let $\mathbf{E}C(4\sqrt{2}|Z|) < +\infty$, let $\sum_{k=1}^{\infty} k^k \mathbf{E}X_1^{2k} < +\infty$, and let* (10) *be satisfied for all $p > 2$, then $T_c(\mu_n, \gamma) \to 0$.*

4 Proofs

Proof of Theorem 1. We first show that Π is a tight set. Indeed, for any positive ε there exist compact sets $K_1, K_2 \in \mathcal{B}(M)$, such that $\mu(K_1) \geq 1 - \frac{\varepsilon}{2}$ and $\nu(K_2) \geq 1 - \frac{\varepsilon}{2}$. Let $\pi \in \Pi$ and let (X, Y) be a random vector with law π. Then,

$$\begin{aligned}\pi(K_1 \times K_2) &= P(X \in K_1, Y \in K_2) = P(X \in K_1) + P(Y \in K_2) - P((X \in K_1) \cup (Y \in K_2)) \\ &\geq \mu(K_1) + \nu(K_2) - 1 \geq (1 - \varepsilon/2) + (1 - \varepsilon/2) - 1 = 1 - \varepsilon.\end{aligned} \tag{12}$$

Since (12) holds for all $\pi \in \Pi$ with the same compact set $K_1 \times K_2$, this proves that Π is tight. Therefore, according to Prokhorov's theorem (Billingsley [3], Section 5), Π is relatively compact.

If $T_c(\mu, \nu) = +\infty$, then $\int c(x, y) d\pi(x, y) = +\infty$, for all $\pi \in \Pi$ and π^* can be chosen to be any probability measure from Π.

If $T_c(\mu, \nu) < +\infty$, then there exists a sequence π_n from Π such that

$$\int c(x, y) d\pi_n(x, y) \to T_c(\mu, \nu). \tag{13}$$

On the other hand, the relative compactness of Π implies the existence of a subsequence π_{n_k} which converges weakly to some probability measure π on $\mathcal{B}(M \times M)$. Let us verify that π is the measure π^* we are looking for. First, we want to prove that $\pi \in \Pi$, i.e. that the marginal distributions of π are μ and ν, respectively.

Let μ_1 and ν_1 be marginals of π. We will check that $\mu_1(B) = \mu(B)$, for any $B \in \mathcal{B}(M)$ such that $\mu_1(\partial B) = 0$. Indeed, since $\partial(B \times M) \subset (\partial B \times M) \cup (B \times \partial M) = \partial B \times M$ (Billingsley [3], (2.8)), we have

$$\pi(\partial(B \times M)) \leq \pi(\partial B \times M) = \mu_1(\partial B) = 0.$$

Therefore, the weak convergence $\pi_{n_k} \Rightarrow \pi$ implies that $\pi_{n_k}(B \times M) \to \pi(B \times M)$, and we obtain

$$\mu(B) = \pi_{n_k}(B \times M) \to \pi(B \times M) = \mu_1(B).$$

Similarly, we can show that $\nu_1(B) = \nu(B)$, for any $B \in \mathcal{B}(M)$ such that $\nu_1(\partial B) = 0$. Finally, it remains to check that two probability measures μ_1 and μ (respectively ν_1 and ν) are the same if they coincide on the Borel sets having a boundary of μ_1-measure (respectively ν_1-measure) zero.

Let $D \in \mathcal{B}(M)$ be a closed set. For $\varepsilon > 0$, let $D^\varepsilon = \{x \in M : d(x, D) < \varepsilon\}$ and let $\mathcal{D} = \{D^\varepsilon, 0 < \varepsilon < 1\}$. Then there exists at most a countable number of ε_k, $0 < \varepsilon_k < 1$, such that sets D^{ε_k} have a boundary of positive μ_1-measure. We remove the sets D^{ε_k} from $\mathcal{D}$, and obtain

$$\mathcal{D}^0 = \{D^\varepsilon, 0 < \varepsilon < 1, \mu_1(\partial D^\varepsilon) = 0\}.$$

We can then choose a decreasing sequence $\varepsilon_n \to 0$, $0 < \varepsilon_n < 1$, with $D_n = D^{\varepsilon_n} \in \mathcal{D}^0$. The sets D_n are such that: (a) $D_{n+1} \subset D_n$ for all n; (b) $\bigcap_n D_n = D \cup \partial D = D$; (c) $\mu_1(D_n) = \mu(D_n)$. The properties (a)–(c) yield

$$\mu_1(D) = \mu_1(\bigcap_n D_n) = \lim_{n \to \infty} \mu_1(D_n) = \lim_{n \to \infty} \mu(D_n) = \mu(D).$$

Therefore, the measures μ_1 and μ coincide on all the closed subsets of M. Since $\mathcal{B}(M)$ is generated by such sets, we conclude that $\mu_1 = \mu$. Similar arguments lead to $\nu_1 = \nu$. We have proved that the probability measure π has respective marginals μ and ν, i.e. $\pi \in \Pi$.

Next, we will check that $\int c(x, y) d\pi(x, y) = T_c(\mu, \nu)$. Since c is lower semicontinuous, for any real b the set $\{(x, y) : c(x, y) > b\}$ is open ([3], Appendix I). Let $A = \{(x, y) : c(x, y) > 0\}$.

Then the weak convergence $\pi_{n_k} \Longrightarrow \pi$ and (13) imply that

$$\begin{aligned}\int c(x,y)d\pi(x,y) &= \int_A c(x,y)d\pi(x,y)\\ &\leq \lim\ \inf_{n_k} \int_A c(x,y)d\pi_{n_k}(x,y)\\ &= \lim\ \inf_{n_k} \int c(x,y)d\pi_{n_k}(x,y)\\ &= T_c(\mu,\nu).\end{aligned}$$

Since $\pi \in \Pi$, the reverse inequality $\int c(x,y)d\pi(x,y) \geq T_c(\mu,\nu)$ holds true. We thus conclude that $\int c(x,y)d\pi(x,y) = T_c(\mu,\nu)$. In other words, the transportation distance becomes the total transportation cost associated to the measure π. Finally, we set $\pi^* = \pi$ and the proof is now complete. $\square$

Proof of Theorem 2 and Corollary 1. Assume that both (a) and (b) are satisfied. Let X, X_n be random elements with respective distributions μ and μ_n and such that X and X_n are independent, for any n. Then $(C(2d(X_n,a)))$ is uniformly bounded, that is $I_1 = \sup_n \mathbf{E}C(2d(X_n,a)) < \infty$. Set $I_2 = \mathbf{E}C(2d(X,a)) < \infty$.

Fix $\varepsilon > 0$ and choose a compact set K_1 in $\mathcal{B}(M)$ such that $\mu(\partial K_1) = 0$ and

$$\int_{(K_1)^c} C(2d(x,a))d\mu(x) < \varepsilon.$$

The weak convergence $\mu_n \Longrightarrow \mu$ implies the tightness of the family $(\mu_n,\mu)_{n\geq 1}$, thus there exists a compact set $K_2 \in \mathcal{B}(M)$ such that $\mu_n(K_2)^c < \varepsilon$, $\mu(K_2)^c < \varepsilon$ and $\mu(\partial K_2) = 0$. Let $K = K_1 \cup K_2$. Then K is compact, and

$$\int_{K^c} C(2d(x,a))d\mu(x) < \varepsilon, \tag{14}$$

$$\mu_n(K^c) < \varepsilon, \quad \mu(K^c) < \varepsilon, \tag{15}$$

with also $\mu(\partial K) = 0$, since $\mu(\partial K) \leq \mu(\partial K_1) + \mu(\partial K_2)$. Since (b) holds, we can choose a positive integer N_1 such that for any $n \geq N_1$,

$$\left|\int C(2d(x,a))d\mu_n(x) - \int C(2d(x,a))d\mu(x)\right| < \varepsilon. \tag{16}$$

As $X_n \xrightarrow{d} X$, for the chosen compact set K and the continuous function $C(2d(\cdot,a))$ we have

$$\mathbf{E}C(2d(X_n,a))\mathbf{1}_{\{X_n\in K\}} \to \mathbf{E}C(2d(X,a))\mathbf{1}_{\{X\in K\}}.$$

Hence, we can choose a positive integer N_2 such that, for any $n \geq N_2$,

$$\left|\int_K C(2d(x,a))d\mu_n(x) - \int_K C(2d(x,a))d\mu(x)\right| < \varepsilon. \tag{17}$$

Then for $n \geq \max\{N_1, N_2\}$, the estimates (14), (16) and (17) yield

$$\begin{aligned}
&\left|\int_{K^c} C(2d(x,a))d\mu_n(x)\right| \\
&\qquad \leq \left|\int C(2d(x,a))d\mu_n(x) - \int C(2d(x,a))d\mu(x)\right| \\
&\qquad + \left|\int_K C(2d(x,a))d\mu_n(x) - \int_K C(2d(x,a))d\mu(x)\right| + \left|\int_{K^c} C(2d(x,a))d\mu(x)\right| \\
&\qquad < 3\varepsilon.
\end{aligned} \tag{18}$$

The weak convergence $X_n \xrightarrow{d} X$ implies that $C(2d(X_n, X))\mathbf{1}_{\{X_n \in K,\ X \in K\}} \xrightarrow{d} 0$. The continuous function $C(d(x,y))$ is bounded on the compact set $K \times K$, therefore

$$\mathbf{E}C(d(X_n, X))\mathbf{1}_{\{X_n \in K,\ X \in K\}} \to 0.$$

This means that there exists a positive integer N_3 such that, for any $n \geq N_3$,

$$\left|\int_K \int_K C(d(X_n, X))d\pi_n(x,y)\right| < \varepsilon, \tag{19}$$

where π_n is the joint distribution of X_n and X.

Since C is a non-negative and non-decreasing,

$$\begin{aligned}
C(d(x,y)) &\leq C(d(x,a) + d(y,a)) \leq C(2\max\{d(x,a), d(y,a)\}) \\
&\leq C(2d(x,a)) + C(2d(y,a)),
\end{aligned} \tag{20}$$

for all $x, y \in M$.

Using the inequalities (14), (15), (18), (20), and the independence of X_n and X, we have:

$$\begin{aligned}
&\int_{K^c}\int_{K^c} C(d(x,y))d\pi_n(x,y) = \mathbf{E}C(d(X_n, X))\mathbf{1}_{\{X_n \in K^c,\ X \in K^c\}} \\
&\quad \leq \mathbf{E}C(2d(X_n, a))\mathbf{1}_{\{X_n \in K^c\}}\mathbf{1}_{\{X \in K^c\}} + \mathbf{E}C(2d(X, a))\mathbf{1}_{\{X \in K^c\}}\mathbf{1}_{\{X_n \in K^c\}} \\
&\quad \leq \left(\int_{K^c} C(2d(x,a))d\mu_n(x)\right)\mu(K^c) + \left(\int_{K^c} C(2d(x,a))d\mu(x)\right)\mu_n(K^c) \\
&\quad < 3\varepsilon^2 + \varepsilon^2,
\end{aligned} \tag{21}$$

for all $n \geq \max\{N_1, N_2\}$. Similarly,

$$\begin{aligned}
&\int_K \int_{K^c} C(d(x,y))d\pi_n(x,y) = \mathbf{E}C(d(X_n, X))\mathbf{1}_{\{X_n \in K,\ X \in K^c\}} \\
&\quad \leq \mathbf{E}C(2d(X_n, a))\mathbf{1}_{\{X_n \in K\}}\mathbf{1}_{\{X \in K^c\}} + \mathbf{E}C(2d(X, a))\mathbf{1}_{\{X \in K^c\}}\mathbf{1}_{\{X_n \in K\}} \\
&\quad \leq I_1\mu(K^c) + \varepsilon\mu_n(K) \\
&\quad < I_1\varepsilon + \varepsilon,
\end{aligned} \tag{22}$$

and

$$\begin{aligned}
&\int_{K^c}\int_K C(d(x,y))d\pi_n(x,y) = \mathbf{E}C(d(X_n, X))\mathbf{1}_{\{X_n \in K^c,\ X \in K\}} \\
&\quad \leq \mathbf{E}C(2d(X_n, a))\mathbf{1}_{\{X_n \in K^c\}}\mathbf{1}_{\{X \in K\}} + \mathbf{E}C(2d(X, a))\mathbf{1}_{\{X \in K\}}\mathbf{1}_{\{X_n \in K^c\}} \\
&\quad \leq 3\varepsilon\mu(K) + I_2\mu_n(K^c) < 3\varepsilon + I_2\varepsilon,
\end{aligned} \tag{23}$$

for $n \geq \max\{N_1, N_2\}$.

Thus for $n \geq \max\{N_1, N_2, N_3\}$ the inequalities (19), (21)–(23) yield

$$\begin{aligned}
T_c(\mu_n, \mu) \leq \mathbf{E}C(d(X_n, X)) &= \int\int C(d(x,y))d\pi_n(x,y) \\
&= \int_K\int_K C(d(x,y))d\pi_n(x,y) + \int_{K^c}\int_{K^c} C(d(x,y))d\pi_n(x,y) \\
&+ \int_K\int_{K^c} C(d(x,y))d\pi_n(x,y) + \int_{K^c}\int_K C(d(x,y))d\pi_n(x,y) \\
&\leq \varepsilon(6 + 4\varepsilon + I_1 + I_2).
\end{aligned}$$

We conclude that (a) and (b) imply $T_c(\mu_n, \mu) \to 0$.

Next, we assume that $T_c(\mu_n, \mu) \to 0$ and verify that (a) $\mu_n \Longrightarrow \mu$ takes place. According to Theorem 1, for any n there exists a pair of random elements X_n and X with distributions μ_n and μ, respectively, which are minimizers of the total transportation cost: $T_c(\mu_n, \mu) = \mathbf{E}C(d(X_n, X))$. Let us note that X may depend on n, so each time it appears in this proof, we assume that $X = X^{(n)}$. (Of course all the $X^{(n)}$ have the same law μ.)

Since C is a non-negative function, $\mathbf{E}C(d(X_n, X)) \to 0$, that is $C(d(X_n, X)) \xrightarrow{L_1} 0$. This implies that

$$C(d(X_n, X)) \xrightarrow{P} 0. \tag{24}$$

Fix $\varepsilon > 0$. As C is non-decreasing, we have

$$\{d(X_n, X) > \varepsilon\} \subset \{C(d(X_n, X)) \geq C(\varepsilon)\}.$$

The convergence result (24) means that the probability of the last event tends to 0, for any positive $C(\varepsilon)$, as $n \to \infty$. Hence for any $\varepsilon > 0$, $P(d(X_n, X) > \varepsilon) \to 0$, as $n \to \infty$. The convergence, in probability, of $d(X_n, X) = d(X_n, X^{(n)}) \xrightarrow{P} 0$ implies that $\mu_n \Longrightarrow \mu$ (Billingsley [3], theorem 4.1).

Now, assume that the doubling condition (6) is satisfied and let us verify that $T_c(\mu_n, \mu) \to 0$ implies (b$'$). Since $\mu_n \Longrightarrow \mu$ and since $C(d(\cdot, a))$ is continuous on M, weak convergence holds: $C(d(X_n, a) \xrightarrow{d} C(d(X, a))$. In order to verify (b$'$), it thus suffices to check that the sequence $(C(d(X_n, a)))$ is uniformly integrable. The uniform integrability is equivalent to the pair of conditions: (i) $(\mathbf{E}C(d(X_n, a)))$ is uniformly bounded and (ii) for $A \in \mathcal{F}$, $(\mathbf{E}C(d(X_n, a))\mathbf{1}_A)$ is uniformly continuous, (i.e. $\sup_n \mathbf{E}C(d(X_n, a))\mathbf{1}_A \to 0$ as $P(A) \to 0$).

Together (6) and (20) yield the inequalities

$$C(d(x, a)) \leq \lambda C\left(\frac{1}{2}d(x, a)\right) \leq \lambda C(d(x, y)) + \lambda C(d(y, a)), \tag{25}$$

for all $x, y \in M$ and the positive constant λ. Then

$$\mathbf{E}C(d(X_n, X)) \geq \frac{1}{\lambda}\mathbf{E}C(d(X_n, a)) - \mathbf{E}C(d(X, a)). \tag{26}$$

Suppose that $(\mathbf{E}C(d(X_n, a)))$ is not uniformly bounded. Then there exists a subsequence $(\mathbf{E}C(d(X_{n'}, a)))$ such that $\mathbf{E}C(d(X_{n'}, a)) \to +\infty$. Applying (26) to this subsequence, we come to the following contradiction:

$$0 \leftarrow \mathbf{E}C(d(X_n, X)) \geq \frac{1}{\lambda}\mathbf{E}C(d(X_{n'}, a)) - \mathbf{E}C(d(X, a)) \to +\infty.$$

Thus, $(\mathbf{E}C(d(X_n,a)))$ is uniformly bounded.

Let ε be fixed, and let $A \in \mathcal{F}$. Since $T_c(\mu_n,\mu) \to 0$, we can choose a positive integer N such that $\mathbf{E}C(d(X_n,X))\mathbf{1}_A < \varepsilon$, for all $n \geq N$. By applying once again the inequality (26), we obtain

$$\sup_{n\geq N} \mathbf{E}C(d(X_n,a))\mathbf{1}_A \leq \lambda \sup_n \mathbf{E}C(d(X_n,X))\mathbf{1}_A + \lambda \mathbf{E}C(d(X,a))\mathbf{1}_A.$$

Let $P(A) \to 0$. Since $(C(d(X_n,X)))$ is uniformly integrable and since $\mathbf{E}C(d(X,a)) \leq \mathbf{E}C(2d(X,a)) < \infty$,

$$\sup_{n\geq N} \mathbf{E}C(d(X_n,a))\mathbf{1}_A \to 0,$$

i.e. $(\mathbf{E}C(d(X_n,a))\mathbf{1}_A)$ is uniformly continuous. Hence, the sequence $(C(d(X_n,a)))$ is uniformly integrable and (b′) $\int C(d(x,a))d\mu_n \to \int C(d(x,a))d\mu$ holds.

Note that from (6) and since C is non-decreasing, the following two inequalities hold true

$$C(2d(x,a)) \leq \lambda C(d(x,a)), \qquad C(d(x,a)) \leq C(2d(x,a)),$$

for any $x \in M$. This implies that
$(\int C(d(x,a))\mu_n(dx) < \infty) \iff (\int C(2d(x,a))\mu_n(dx) < \infty)$ and that $(\int C(d(x,a))\mu(dx) < \infty) \iff (\int C(2d(x,a))\mu(dx) < \infty)$. Therefore, in the setting of the theorem, the sequences $(C(d(X_n,a)))$ and $(C(2d(X_n,a)))$ are both either uniformly integrable or not, and (b)$\iff$(b′).

This observation completes the proof of Theorem 2 and of Corollary 1. □

Proof of Corollary 2. Let K_1 and K_2 be the respective supports of μ and ν. If μ and ν are absolutely continuous with respective densities f_1 and f_2, then

$$\begin{aligned}\left|\int \phi(x)d\mu - \int \phi(x)d\nu\right| &= \left|\int \phi(x)f_1(x)dx - \int \phi(x)f_2(x)dx\right| \\ &\leq \int_{K_1\cup K_2} |\phi(x)||f_1(x)-f_2(x)|dx \\ &\leq L_\phi \|\mu-\nu\|_{TV},\end{aligned}$$

where $L_\phi = \sup\{|\phi(x)| : x \in \overline{(K_1\cup K_2)}\}$ (here $\overline{A} = A \cup \partial A$).

To prove the result in the general case, define the partition $(A_m)_{m\in\mathbf{Z}}$ of M, $A_m \in \mathcal{B}(M)$, as follows:

$$A_m = \{x \in M : m-1 \leq \phi(x) < m\}.$$

Thus,

$$\begin{aligned}\left|\int \phi(x)d\mu - \int \phi(x)d\nu\right| &= \left|\sum_{m=-\infty}^{+\infty} \int \phi(x)\mathbf{1}_{A_m} d(\mu-\nu)\right| \\ &\leq \sum_{m=-\infty}^{+\infty} |m||\mu(A_m)-\nu(A_m)| \\ &\leq L_\phi \|\mu-\nu\|_{TV},\end{aligned} \tag{27}$$

where L_ϕ is defined as above, and where we used the dual definition of the total variation distance.

Let μ_n and μ be probability measures on M with bounded supports respectively denoted K_n and K. Let also $\cup_n K_n$ be bounded and $\|\mu_n - \mu\|_{TV} \to 0$. Convergence in total variation implies weak convergence $\mu_n \Longrightarrow \mu$. All the conditions of Theorem 2 are satisfied. Therefore, to prove the convergence of μ_n to μ in T_c, it suffices to check that $\int C(2d(x,a))d\mu_n \to \int C(2d(x,a))d\mu$ for some $a \in M$. The inequality (27) yields for any n

$$\left|\int \phi(x)d\mu_n - \int \phi(x)d\mu\right| \le L_\phi \|\mu_n - \mu\|_{TV}, \tag{28}$$

where $L_\phi = \sup\{|\phi(x)| : x \in \overline{\cup K_n}\} < \infty$ does not depend on n. By fixing $a \in M$ and applying (28) to $\phi(x) = C(2d(x,a))$, we obtain that the convergence in total variation implies the convergence of the integrals $\int \phi(x)d\mu_n \to \int \phi(x)d\mu$. This completes the proof. $\square$

Proof of Theorem 3. (i) First, we set $Y_n = S_n/(\sigma\sqrt{n})$ and check that $\mathbf{E}|Y_n|^p \to \mathbf{E}|Z|^p$. The classical CLT gives $Y_n \xrightarrow{d} Z$, while the uniform boundedness of $\mathbf{E}|Y_n|^p$ follows from Rosenthal's inequality (7):

$$\mathbf{E}|Y_n|^p \le K(p)\left(\frac{\mathbf{E}|X_1|^p}{\sigma^p n^{\frac{p}{2}-1}} + 1\right) \le K(p)\left(\frac{\mathbf{E}|X_1|^p}{\sigma^p} + 1\right) \tag{29}$$

Let $Z \sim N(0,1)$ and sequence (X_n) be independent. We fix $\varepsilon > 0$ and choose a compact set K, $K \in \mathcal{B}(\mathbf{R})$, such that

$$\gamma(K^c) < \varepsilon, \quad \mathbf{E}|Z|^p \mathbf{1}_{\{Z\in K^c\}} < \varepsilon, \quad \gamma(\partial K) = 0. \tag{30}$$

Hence,

$$\begin{aligned}|\mathbf{E}|Y_n|^p - \mathbf{E}|Z|^p| &\le \left|\mathbf{E}|Y_n|^p \mathbf{1}_{\{Z\in K\}} - \mathbf{E}|Z|^p \mathbf{1}_{\{Z\in K\}}\right| + \mathbf{E}|Y_n|^p \mathbf{1}_{\{Z\in K^c\}} + \mathbf{E}|Z|^p \mathbf{1}_{\{Z\in K^c\}} \\ &\le \varepsilon + \varepsilon K(p)\left(\frac{\mathbf{E}|X_1|^p}{\sigma^p} + 1\right) + \varepsilon,\end{aligned} \tag{31}$$

for sufficiently large n, thanks to (29), (30) and to the convergence of $\mathbf{E}|Y_n|^p \mathbf{1}_{\{Z\in K\}}$ to $\mathbf{E}|Z|^p \mathbf{1}_{\{Z\in K\}}$. Therefore, $\mathbf{E}|Y_n|^p \to \mathbf{E}|Z|^p$. This, in particular, implies the uniform integrability of $(|Y_n|^p)$.

Next, we show that for a cost function with $C(x) = O(x^p)$, at infinity, all the conditions of Theorem 2 are satisfied. We have $\mathbf{E}C(2|Z|) < +\infty$, while the finiteness of $\int C(2|x|)\mu_n(dx)$ follows from (29) and inequality

$$C(2|Y_n|) = C(2|Y_n|)\mathbf{1}_{\{|Y_n|\le x_0\}} + C(2|Y_n|)\mathbf{1}_{\{|Y_n|>x_0\}} \le C(x_0)\mathbf{1}_{\{|Y_n|\le x_0\}} + \beta|Y_n|^p \mathbf{1}_{\{|Y_n|>x_0\}}, \tag{32}$$

with a positive constant β such that $C(x) \le \beta x^p$, for all $x > x_0$. The CLT provides the weak convergence $\mu_n \Longrightarrow \gamma$, while we obtain $\mathbf{E}C(2|Y_n|) \to \mathbf{E}C(2|Z|)$ from the uniform integrability of $(C(2|Y_n|)$ which follows from (32) and uniform integrability of $(|Y_n|^p)$. Thus, applying Theorem 2, we obtain that $T_c(\mu_n, \gamma) \to 0$.

Remark 2. Since $\mathbf{E}|Y_n|^p \to \mathbf{E}|Z|^p$, the convergence of $\mathbf{E}C(|Y_n|)$ to $\mathbf{E}C(|Z|)$ follows from part (a) and part (c) of the result of Bickel and Freedman [2] cited above. Then, $\mathbf{E}C(2|Y_n|) \to \mathbf{E}C(2|Z|)$ is implied by the doubling condition (6).

(ii) Once again, we check that the conditions of Theorem 2 are satisfied. First, $\mathbf{E}C(4\sqrt{2}|Z|) < \infty$ implies that $\mathbf{E}C(2|Z|) < \infty$ and that $C(2x) = o(e^{x^2/16})$.

The function $f(x) = \exp(x^2/16)$ has the expansion

$$f(x) = 1 + \sum_{k=1}^{\infty} \frac{(2k-1)!!}{2^{3k}(2k)!} x^{2k}. \tag{33}$$

Then Stirling's formula yields, for a generic term f_k of series (33), starting at some index k_0

$$f_k = \frac{e^{k+1}(k-\frac{1}{2})^{2k-\frac{5}{2}}}{2^{4k+1}(k-2)^{k-\frac{3}{2}}k^{2k+\frac{1}{2}}} x^{2k}. \tag{34}$$

Next, the Rosenthal inequality (7) can be written, with constants, in the following form (Petrov [17], inequality (2.35)):

$$\mathbf{E}|S_n|^k \le k^k n \mathbf{E}|X_1|^k + \frac{4k^{\frac{k}{2}+1}}{2^k} e^k n^{\frac{k}{2}} \sigma^k. \tag{35}$$

Together (33)–(35) imply, for $Y_n = S_n/(\sigma\sqrt{n})$,

$$\begin{aligned}
\mathbf{E}f(|Y_n|) &\le M + \sum_{k=k_0}^{\infty} \frac{e^{k+1}(k-\frac{1}{2})^{2k-\frac{5}{2}}\mathbf{E}|X_1|^{2k}}{2^{2k+1}\sqrt{k}(k-2)^{k-\frac{3}{2}}\sigma^{2k}n^{k-1}} + \sum_{k=k_0}^{\infty} \frac{e^{3k+1}(k-\frac{1}{2})^{2k-\frac{5}{2}}}{2^{5k}\sqrt{k}(k-2)^{k-\frac{3}{2}}k^{k-\frac{1}{2}}} \\
&\le M + \beta_1 \sum_{k=k_0}^{\infty} k^2 \mathbf{E}|X_1|^{2k} + \beta_2 \sum_{k=k_0}^{\infty} \left(\frac{e^3}{2^5}\right)^k \\
&= M + \beta_1 Q_1 + \beta_2 Q_2 < +\infty
\end{aligned} \tag{36}$$

for some positive constants M, β_1, and β_2.

Since $C(2x) = o(f(x))$, there exists $x_0 > 0$ such that $C(2x) \le f(x)$, for all $x > x_0$. The inequality (36) gives

$$\begin{aligned}
\mathbf{E}C(2|Y_n|) &= \mathbf{E}C(2|Y_n|)\mathbf{1}_{\{|Y_n| \le x_0\}} + \mathbf{E}C(2|Y_n|)\mathbf{1}_{\{|Y_n| > x_0\}} \\
&\le C(2x_0) + M + \beta_1 Q_1 + \beta_2 Q_2.
\end{aligned} \tag{37}$$

Therefore, $(\mathbf{E}C(2|Y_n|))$ is bounded and, moreover, uniformly bounded. Next, we check that $(\mathbf{E}C(2|Y_n|)\mathbf{1}_A)$, $A \in \mathcal{F}$, is uniformly continuous.

Fix $\varepsilon > 0$, and choose x_1 positive and such that $C(2x) \le \varepsilon_1 f(x)$, for all $x > x_1$, with $\varepsilon_1 = \frac{\varepsilon}{2(M+\beta_1 Q_1+\beta_2 Q_2)}$. If $P(A) = \frac{\varepsilon}{2C(2x_1)}$, and in complete similarity to (37) we have

$$\sup_n \mathbf{E}C(2|Y_n|)\mathbf{1}_A \le C(2x_1)P(A) + \varepsilon_1 \sup_n \mathbf{E}f(|Y_n|)\mathbf{1}_A < \varepsilon. \tag{38}$$

The inequalities (37) and (38) yield the uniform integrability of $(C(2|Y_n|))$, while the classical CLT provides $\mu_n \Longrightarrow \gamma$. Hence, $\mathbf{E}C(2|Y_n|) \to \mathbf{E}C(2|Z|)$ and all the conditions of Theorem 2 are satisfied. Then $T_c(\mu_n, \gamma) \to 0$. This concludes the proof of Theorem 3. $\square$

Proof of Theorems 4 and 5. They are carried out by using the same arguments as in the proof of Theorem 3.

For sequences of dependent random variables we assume additional conditions of asymptotic independence, which yield the moment inequalities (9) and (11) for the moments of the sums: the condition (8) for strongly mixing sequences and the condition (10) for associated sequences. We also use the bounds on $K(p)$ in (9) and (11) derived by Doukhan and Louhichi in [10].

References

[1] Ambrosio, L., Gigli, N. and Savaré, G. (2005) *Gradient Flows: in Metric Spaces and in the Space of Probability Measures.* Birkhäuser, Boston.

[2] Bickel, P.J. and Freedman, D.A. (1981) *Some asymptotic theory for the bootstrap.* Ann. Stat., 9, pp. 1196–1217.

[3] Billingsley P. (1968) *Convergence of Probability Measures.* Wiley, New York.

[4] Birkel, T. (1988) *Moment bounds for associated sequences.* Ann. Probab., 16, 3, pp. 1184–1193.

[5] Caffarelli, L. (1996) *Allocation maps with general cost functions.* Partial Differential Equations and Applications, Lecture Notes in Pure and Applied Math., 177, pp. 29–35.

[6] Cuesta, J.A. and Matran, C. (1989) *Notes on the Wasserstein metric in Hilbert spaces.* Ann. Probab., 17, 3, pp.1264–1276.

[7] Dall'Aglio, G. (1956) *Sugli estremi dei momenti delle funzioni di ripartizione doppia.* Ann. Scuola Normale Superiore Di Pisa, Cl. Sci. 3, 1, pp. 33–74.

[8] Denker, M. (1986) *Uniform integrability and the central limit theorem for strongly mixing processes.* Dependence in Probability and Statistics, (E. Eberlein and M.S. Taqqu, eds.) Progress in Probability and Statistics, 11, pp. 269–274. Birkhäuser, Boston

[9] Dobrushin, R.L. (1970) *Prescribing a system of random variables by conditional distributions.* Theor. Prob. Appl., 15, pp. 458–486.

[10] Doukhan, P. and Louhichi, S. (1999) *A new weak dependence condition and applications to moment inequalities.* Stochastic process. Appl, 84, 2, pp. 313–342.

[11] Gangbo, W. and McCann, R.J. (1996) *The geometry of optimal transportation.* Acta Math., 177, pp. 113–161.

[12] Johnson, O. and Samworth, R. (2005) *Central limit theorem and convergence to stable laws in Mallows distance.* Bernoulli, 11, 5, pp. 829–845.

[13] Johnson, O. and Samworth, R. (2006) *Acknowledgement of priority: Central limit theorem and convergence to stable laws in Mallows distance.* Bernoulli, 12, 1, p. 191.

[14] Major, P. (1978) *On the invariance principle for sums of independent identically distributed random variables.* J. Mult. Anal., 8, pp. 487–517.

[15] Mallows, C.L. (1972) *A note on asymptotic joint normality.* Ann. Math. Stat., 43, 2, pp. 508–515.

[16] Newman, C.M. (1980) *Normal fluctuations and the FKG-inequalities.* Commun. Math. Phys., 74, pp. 119–128.

[17] Petrov, V.V. (1995) *Limit Theorems of Probability Theory: Sequences of Independent Random Variables* Clarendon Press, Oxford.

[18] Rachev, S.T. (1984) *On a class of minimal functionals on a space of probability measures.* Theory of Probability and its Applications 29, 1, pp. 41–49.

[19] Rachev, S.T. and Rüschendorf, L. (1998) *Mass Transportation Problems.* Vol. I, II. Springer-Verlag, New York.

[20] Salvemini, T. (1943) *Sul calcolo degli indici di concordanza tra due caratteri quantitativi.* Atti della VI Riunione della Soc. Ital. di Statistica, Roma.

[21] Tanaka, H. (1973) *An inequality for a functional of probability distributions and its application to Kac's one dimensional model of a Maxwellian gas.* Z. Wahrscheinlichkeitstheorie Verw. Geb., 27, pp.47–52.

[22] Yokoyama, R. (1980) *Moment bounds for stationary mixing sequences.* Z. Wahrsch. Verw. Gebiete, 52, pp. 45–57.